chinaXiv:201908.00074v1

计算精神病学：抑郁症研究和临床应用的新视角

区健新[1]；吴 寅[1]；刘金婷[1]；李 红[1,2,3]

([1]深圳大学心理学院; [2]深圳市情绪与社会认知科学重点实验室;
[3]深圳市神经科学研究院，深圳, 518060)

摘 要 抑郁症是一种复杂而异质的精神疾病，给全球带来沉重的疾病负担。尽管基于症状学的诊断方法已被广泛应用于各领域，但这种方法并不利于病理机制的探讨。另外，该诊断方法预测效度较低，导致其难以准确评估和比较各种治疗方案的疗效。计算精神病学方法则能通过理论驱动和数据驱动两种互补的方法解决上述问题，从而提高对抑郁症的认识、预防和治疗。理论驱动方法基于经验知识或假设，利用计算建模方法对数据进行多水平分析；数据驱动方法则基于机器学习算法分析高维数据，提高抑郁症诊断和预测的准确性，进而提高治疗的精准度。理论驱动和数据驱动方法的发展与结合，以及人才和资源的整合，将会更有效地推进抑郁症的防治。

关键词 抑郁症；计算精神病学；计算模型；机器学习；诊断；治疗

抑郁症是一种常见且容易被忽视的精神疾病，它已成为全球人“精神残疾”的主要因素，尤其是中青年群体(Vos et al., 2016)。2018 年 3 月，世界卫生组织(World Health Organization, WHO)公布，全球约有 3 亿人遭受抑郁症的困扰，每年约有 80 万人因抑郁而自杀身亡(WHO, 2018)。抑郁症大大增加了全球疾病的总体负担，它给病患及其家属带来极大的痛苦(Herrman et al., 2019; Ledford, 2014)。因此，对抑郁症进行系统深入的研究，提出有效的预防和治疗方案刻不容缓。然而，基于疾病症状学的抑郁症诊断方法，在临床应用中效果不佳，预测效度较低(Cuthbert & Insel, 2013)，后续发展出来的生物精神病学(biological psychiatry)也未能较好地解释抑郁症的心理病理机制。随着计算机科学、生物技术等的发展，新诞生的计算精神病学(computational psychiatry)力图通过引入计算和统计的方法来打开抑郁症的黑箱，从理论驱动和数据驱动两方面，揭示大脑的信息加工过程和精神疾病的发病机制，并利用高维的复杂数据来鉴别精神疾病(Huys, 2018a)。计算精神病学是计算神经科学和精神病学交叉融合的前沿领域，它将对精神疾病的发病机制、诊断、治疗和预防产生巨大的推动作用。

1 什么是抑郁症

1.1 抑郁症的定义

抑郁通常指暂时的情绪低落、悲伤等情绪，而抑郁症则是一种涉及基因、神经系统和认知等多层面因素的精神疾病。抑郁障碍(depressive disorder)是一种心境障碍，通常表现为情绪低落、快感缺失、内疚感和疲惫感强、自我价值感低，并伴随着睡眠障碍、食欲减

收稿日期：2019-01-29
国家自然科学基金（31671150）、广东省普通高校创新团队建设项目（2015KCXTD009）、广东省省级基础研究及应用研究重大项目（2016KZDXM009）、深圳市基础研究布局项目（JCYJ20150729104249783）、深圳市孔雀计划项目（KQTD2015033016104926）、深圳市基础研究布局项目（JCYJ20170818110103216, JCYJ20170412164413575）、国家自然科学基金（31600928）
通信作者：李红，Email: lihongszu@szu.edu.cn

退、注意力不集中等症状(Lu et al., 2014)。研究发现，抑郁群体通常表现出负性认知偏向和情绪调节不良的现象(Hammen, 2018)，并在趋近性动机行为和奖赏学习等过程上也存在一定的损伤(Cooper, Arulpragasam, & Treadway, 2018)。值得注意的是，抑郁障碍存在很高的异质性。抑郁症可分为多种亚型，而且共病性(comorbidity)强，几乎三分之二的重性抑郁障碍(major depressive disorder)患者共患焦虑障碍(anxiety disorder) (Goldberg & Fawcett, 2012)；另外，不同患者的抑郁发作时长和次数也不相同(Malhi & Mann, 2018)。然而，上述观点只描述了抑郁症患者的症状表现，并未能解释和控制抑郁症状的发生、发展、维持和消退。

1.2 抑郁症的诊断

目前，抑郁症的诊断主要基于国际疾病分类标准(International Classification of Diseases, ICD)和美国精神病学会(American Psychiatric Association, APA)出版的精神疾病诊断统计手册(Diagnostic and Statistical Manual of Mental Disorders, DSM)。ICD 和 DSM 基于现象学，以症状群定义和诊断精神障碍。以重性抑郁障碍为例，DSM-V 中对其的诊断标准为，两周以内几乎每天的大多数时间出现持续的抑郁情绪、兴趣或快感缺失、失眠或过度睡眠、体重骤降或骤升、疲惫乏力等九大症状的五种及以上(American Psychiatric Association, 2013)。

尽管这种诊断方式广泛用于基础研究和临床应用，但也存在一些不可忽视的问题。首先，该诊断方式基于患者的自我报告，受个人主观信念和自我觉察能力影响；而后续的临床访谈依赖于患者的口头表达能力和临床医师的诊断经验与能力。这种模糊的症状描述不利于临床医师的客观诊断。其次，这种疾病分类法难以解释共病现象(Hammen, 2018)。抑郁症可以被分为多种亚型，而且与其他精神疾病的症状存在一定的相似性，如双相情感障碍(bipolar affective disorder, BD)、创伤后应激障碍(post-traumatic stress disorder, PTSD)、焦虑障碍等。不同疾病之间具有相似症状，可能预示着这些疾病在更底层因素（如内表型; Pizzagalli, 2014）上存在共同的机制。目前的诊断方式单纯地将这些疾病划分为不同类别，这可能忽视了这些疾病背后的共性，不利于研究者对这些疾病的共病机制进行深入的探讨。再次，ICD 或 DSM 对症状的描述是定性的而非定量的，这些方法粗糙地将不同患者在某症状表现上的不同强度直接二分为有无该症状，不能很好地依据不同症状的严重程度进行个体化的干预和治疗。最后，对精神疾病的诊断常用二分法来划分不同的患病群体和健康群体。某些个体虽尚未达到患病的标准，罹患抑郁症的风险却较大，而传统的二分法则忽略了对该群体的干预。

自我报告和临床访谈的主观性、抑郁与其他精神疾病的共病性、症状和患病的二分划分法，都制约了诊断过程的有效性和实用性。鉴于此，生物精神病学旨在从生物层面探讨精神障碍背后神经系统的功能及其机制(Kapur, Phillips, & Insel, 2012)，并尝试以生物标记物(biomarker)作为抑郁症诊断的依据。尽管生物精神病学考察了多层面的生物病理机制，强调了人体内环境的变化，在一定程度上弥补 ICD 和 DSM 的不足。但也有学者指出，生物上的改变与症状之间的因果关系几乎不可能一一对应(Kendler, 2008)，因为心理健康不仅与复杂的大脑功能相联系，还与个体所处的环境及其经历密切相关(Huys, Maia, & Frank, 2016)。正因为精神疾病涉及多系统多层面的影响，具有很强的异质性(Insel & Cuthbert, 2015)，导致生物精神病学仍面临着一个巨大的鸿沟，即在分子水平和临床症状之间缺少一个合适的中介性描述(Montague, Dolan, Friston, & Dayan, 2012)。计算精神病学引入计算建模的方法，为微观层面（分子、细胞等）和宏观层面（行为、环境等）搭建一座桥梁，尝试揭示不同层面和不同系统间的影响路径，比如揭示某一因子水平的变化，如何影响整个系统的改变，进而引发行为上的改变(Maia & Frank, 2011)。

chinaXiv:201908.00074v1

2 计算精神病学

2.1 定义和目标

计算精神病学整合了来自精神病学、心理学、神经科学、行为经济学和机器学习的方法，基于精神疾病相关的神经和认知现象，建立脑功能的计算模型，如神经元网络的具体模型、高水平认知能力的抽象模型等(Huys, 2014)。它希望通过详细的多维计算模型，来预测心理功能的异常程度并评估治疗方案的疗效(Montague et al., 2012)。

2.2 主要方法

2.2.1 理论驱动的方法

理论驱动的方法源于计算神经科学(computational neuroscience)。目前，计算模型已被用于研究个体的信息表征和加工过程(Sejnowski, Koch, & Churchland, 1988)。它能寻求对这些数据的解释，而非统计其特征，并且将“不可观测的大脑状态”至行为测量这一认知和神经生理过程具象化为数学模型(Stephan, Iglesias, Heinzle, & Diaconescu, 2015)。众多模型普遍认为，人的认知方式（信息加工方式）与个体信息的获取紧密相连，它们交互影响着学习和信念或认知图式(schema)的形成，这些信念或图式外显或内隐地影响个体的行为或决策。长期积累的异常学习经验是形成不良认知方式的核心成分，这种不良认知方式可能是诱发精神疾病的重要因素。

以理论驱动的计算建模方法是解释人类大脑计算属性的独特方法。该方法通过表征特定心理过程的模型参数，从客观上定量地解释该成分在抑郁症患者认知和情绪等功能异常上的权重和影响路径，能够克服症状学诊断的主观性和低精度等问题，为精神疾病的严重程度的判定和疗效评估提供客观的量化指标。一些认知任务若具有较高的抑郁症鉴别敏感性与特异性，则可以用于鉴别健康群体和抑郁群体，提高鉴别的客观性和效率，并依据患者特定认知功能的损伤情况提供精准的认知行为干预方案。

2.2.2 数据驱动的方法

研究者通过生物技术和脑成像技术，获得了丰富的分子生物学和神经影像学的数据，这些数据加上症状和自陈量表的结果，形成了一组复杂的多维数据。这类多维数据正好反映了精神疾病成因的复杂性，各层面的因素均不能随意忽略。然而，传统方法却很难统合这类多维数据。机器学习是一种可供选择的方法，它通过训练，不断积累经验、更新算法，来帮助人们分析庞大而复杂的数据集(Libbrecht & Noble, 2015)。目前，机器学习方法已经广泛应用于情绪状态的区分(Wang, Nie, & Lu, 2014)、心境波动的监测(Eldar, Roth, Dayan, & Dolan, 2018)，以及精神疾病的诊断，如抑郁症(Nouretdinov et al., 2011; Patel, Khalaf, & Aizenstein, 2016)和精神分裂症(Honnorat, Dong, Meisenzahl-Lechner, Koutsouleris, & Davatzikos, 2017; Lin et al., 2017)等。因此，数据驱动的方法通过统合不同层面、不同系统的患者数据，客观地揭示数据集之中有效的临床表现模式(Chekroud, 2017)，进而鉴别健康群体和抑郁群体、评估药物和干预的疗效等，也可以预测罹患抑郁症的风险。同时，其结果能为后续治疗和更深入的病理研究提供参考，促进研究者对抑郁症病理理论的建立与更新。

3 理论驱动的计算精神病学研究

基于理论建立的模型主要包含四类：基于生理的模型(biophysically based models)，关注细胞、突触和皮层等不同水平的活动；联结主义和神经模型(connectionist and neural models)，反映神经水平和行为水平；算法模型(algorithmic models)，如强化学习模型(reinforcement learning model, RL)和漂移扩散模型(drift diffusion model, DDM)，以抽象方式模拟大脑的计算和行为执行；标准化模型(normative models)，基于贝叶斯理论，探讨行为和神经活动是否与标准理论一致(Paulus, Huys, & Maia, 2016)。目前，强化学习模型、漂移

chinaXiv:201908.00074v1

扩散模型、基于贝叶斯假设(Bayesian brain hypothesis)的生成模型(generative model)在精神疾病的研究中最为广泛，并且能从行为到神经等不同层面对更细致的生理心理过程（如注意，决策）进行量化，下文主要介绍上述三种模型在抑郁症研究中的结果。

3.1 强化学习模型

强化学习指的是学习的主体(agent)学会将情境与行动联系在一起的过程，即依据奖赏最大化原则，通过一系列尝试–错误(trial-and-error)的探索，执行目标导向性(goal-directed)的行为(Sutton & Barto, 1998)。主体能感知其所处环境的状态，并有能力作出相应的行动对此状态加以改变。对行为的结果而言，行动不仅会影响奖赏的获得，还会引起后续情境对奖赏的期待(Sutton & Barto, 2018)。

强化学习理论认为，习惯性行为的价值是通过尝试–错误的体验形成的，并未涉及外显决策模型的编码，执行迅速，通常被认为是无模型的强化学习(model-free reinforcement learning)；而目标导向性行为需要审议性规划(deliberative planning)，则被认为是基于模型的强化学习(model-based reinforcement learning)，因为实时计算行为的价值需要内部模型的参与(Daw, Niv, & Dayan, 2005)。Daw 等人(2011)和 Wunderlich 等人(2012)的研究、Doll 等人(2016)的基因研究也证明了这两类学习的存在。

在无模型的学习过程中，最简单、使用最普遍的模型是 Rescorla-Wagner 模型(Sutton & Barto, 1998)。基于该模型假定，在当前 t 试次中，大脑计算的价值(value, V_t)可以表示为：

$$V_t = V_{t-1} + \alpha \cdot PE_{t-1}$$

其中，α 表示学习率(learning rate)，PE 为预测误差(prediction Error)。在前一个($t-1$)试次中，PE 等于该试次所获奖赏(reward, R)与所期望奖赏(expected value, V)之间的差值。

$$PE_{t-1} = R_{t-1} - V_{t-1}$$

学习率 α 是无模型强化学习的核心变量，反映的是主体利用预测误差信息来更新价值函数的效率。价值函数是可以灵活更新的，基于主体自身的动机状态，以及对无奖惩环境的认知，利用算法能更新价值函数(Lee, Seo, & Jung, 2012)。基于 Rescorla-Wagner 模型，研究者发展出更深入的时间差分(temporal difference, TD)模型，该模型将未来的奖赏也考虑在内(详见 Sutton & Barto, 1998; Niv, 2009)　。此外，关于强化学习模型的更多算法，详见 Sutton 和 Barto (Sutton & Barto, 2018)的介绍。

经典的强化学习范式是呈现一对无意义刺激让被试选择，选择不同的刺激会产生不同的结果，即其中一个刺激是高概率获得奖赏或惩罚的(70%~80%)，另一个刺激是低概率获得奖赏或惩罚的(20%~30%)。随着试次数目的增加，被试会发现奖赏或惩罚与特定刺激的联结，并倾向于作出趋近奖赏而回避惩罚的行为。该范式在研究抑郁症群体对奖赏与惩罚的加工中最为普遍，下文将以抑郁症群体为例，介绍强化学习模型解释抑郁症群体快感缺失的机制。

抑郁症的快感缺失与奖赏加工异常存在较大联系。Rothkirch 等人(2017)通过上述范式研究了未接受药物治疗的重性抑郁障碍患者对奖赏和惩罚的加工机制。结果发现，患者能根据金钱奖赏的强化来调节自己的行为，且腹侧纹状体(ventral striatum)和前脑岛(anterior insula)加工奖赏和惩罚信息的神经活动并未损伤。但是，患者内侧眶额皮层(medial orbitofrontal cortex)的活动减弱，而且受内侧眶额皮层影响的奖赏预测误差(reward prediction error)信号与快感缺失呈负相关。因此，抑郁症患者可能在体验快乐或奖赏上存在损伤，该现象与内侧眶额皮层加工预测误差信号有关。此外，大量研究指出，奖赏预测误差信号与腹侧被盖/黑质(ventral tegmental area/substantia nigra)的多巴胺及其投射区域有关，如纹状体(striatum)、前额叶(prefrontal cortex)、海马回(hippocampus)等，并且奖赏预测误差信号在抑郁症群体中通常存在减弱的趋势(Chen, Takahashi, Nakagawa, Inoue, & Kusumi, 2015)。然而，有研究采用相似的范式发现，抑郁症群体加工多巴胺能的奖赏预测误差功能是正常的

chinaXiv:201908.00074v1

chinaXiv:201908.00074v1

(Rutledge et al., 2017)。不一致结果的出现可能与抑郁症患者的同质性、实验任务的难度、是否服用抗抑郁药等有关。被试的抑郁类型、抑郁程度不同，以及药物的剂量和疗效不一，会使得组内差异变大，从而掩盖了实验条件的效应。另外，实验任务过于复杂可能使抑郁症患者的表现出现地板效应，这可能由其学习能力不足引起，但也可能由其注意的维持和执行功能的缺陷等造成。因此，考察抑郁症患者的学习过程需要采用精细的实验设计来严谨地进行推论。

与预测误差反映的无模型学习不同，基于模型的学习需要个体进行审议性规划。研究表明，认知功能受损的抑郁患者较少地使用基于模型的学习，相关的神经活动则表现为外侧和内侧前额叶、眶额皮层和海马回的功能紊乱(Clark, Chamberlain, & Sahakian, 2009; Pizzagalli, 2014; Russo & Nestler, 2013)。可能的原因是，过往学习使其形成了情境是不可控的信念，影响了后续探索和使用基于模型的学习策略来更新已有信念的动机(Montague et al., 2012)，这与习得性无助现象较为相似。除了认知和记忆，应激(stress)等情境因素也会影响基于模型的强化学习。应激损伤了前额叶的结构和功能(Arnsten, 2009; McEwen & Morrison, 2013)，使得行为从目标导向转变为习惯性控制(Schwabe, 2013; Schwabe & Wolf, 2011)。慢性应激水平较高的个体在急性应激的影响下表现出更少的基于模型的学习，执行控制能力较弱(Radenbach et al., 2015)。另外，相比于抑郁程度较低的群体，抑郁程度较高的群体在基于模型的学习上受应激的影响更大，更多地表现出习惯性行为，可能的原因是其执行控制功能的损伤(Heller, Ezie, Otto, & Timpano, 2018)。因此，应激可能通过影响人的目标导向性行为的计划、执行与控制，进而加重抑郁症状。

综上所述，基于强化学习模型的抑郁症研究主要发现了抑郁症群体在加工奖赏预测误差和期望值上存在损伤，表现出较低的奖赏敏感性(Chen et al., 2015)。而且，众多研究较为一致地指出，抑郁症群体在基于模型的学习上表现下降，这可能与患者的工作记忆、执行控制功能和价值加工的损伤有关。

3.2 漂移扩散模型

漂移扩散模型是序贯抽样模型(sequential sampling model)之一，主要用于描述快速的二分决策过程。该简单的二分决策持续时间较短，反应时通常小于 1000 或 1500 毫秒，只存在单一的判断过程，并不涉及多阶段的决策或推理。DDM 假设，决策是噪音信息(noisy evidence)随着时间积累，从一个起点(starting point)，最终达到决策阈限(decision threshold)或边界(boundary)，并产生反应的过程(Ratcliff, 1978; Ratcliff & McKoon, 2008; Voss, Nagler, & Lerche, 2013)。该模型的表达式为：

$$dy(t) = v(\Delta u) \cdot dt + \sigma \cdot dW$$

其中，$y(t)$表示的是在时间点 t 时信息积累的数量，Δu指的是两个选择之间效用(utility)的差异。v 为漂移率(the drift rate)，指的是单位时间内信息积累的数量，即信息加工的速度。σ 指的是 dW 维纳过程（Wiener process，连续时间随机过程）中的高斯噪音参数(Gaussian noise parameter)。对于不同的试次，这些参数会发生改变，任一参数的改变将会影响正确和错误反应的反应时分布(Forstmann, Ratcliff, & Wagenmakers, 2016; Pe, Vandekerckhove, & Kuppens, 2013)。例如，漂移率更大的试次中，正确率会更高，反应时更短；漂移率更小的试次则相反。有研究表明，DDM 能有效地对二分知觉决策的结果提供较为精确的定量解释(Gold & Shadlen, 2007; Hanks, Ditterich, & Shadlen, 2006; Mazurek, Roitman, Ditterich, & Shadlen, 2003)，这一特性促使了研究者通过分析不同信息加工过程中出现的个体差异，进而区分出不同的人群，甚至还能区分出不同病人群体的信息加工异常的阶段及其原因。

近年来，许多研究者用 DDM 来分析不同的信息加工过程(Gomez & Perea, 2014)，如视觉(Krajbich & Rangel, 2011)、注意(Dutilh et al., 2012)、情绪(White, Ratcliff, Vasey, &

McKoon, 2009; White, Ratcliff, Vasey, & McKoon, 2010a)。精神疾病的研究也开始采用 DDM 的方法(White, Ratcliff, Vasey, & McKoon, 2010b)。例如，研究者通过使用 DDM，发现了一些关于抑郁症(depression)的认知、行为和生理的标记物(marker)。

Pe 等人(2013)通过情绪 flanker 任务来探究反刍(rumination)倾向、抑郁倾向与注意偏向之间的关系。结果显示，反刍倾向而非抑郁倾向解释了抑郁组对负性信息的注意偏向。另外，Dillon 等人(2015)则利用传统的 flanker 任务探讨了重性抑郁障碍个体的优势反应偏差（即对分心刺激的反应倾向）、反应抑制和执行控制功能，发现重性抑郁障碍组被试比健康组被试漂移率更低，表现为反应更慢，正确率更高。该结果与过往研究一致，但这种差异的产生并非由简单的反应时-正确率权衡(speed-accuracy trade-off)引起，因为重性抑郁障碍组的决策阈限并未改变(Dutilh et al., 2012)。另外，快感缺失得分与漂移率呈负相关。结果表明，重性抑郁障碍个体优势反应偏差降低，执行控制功能减弱，这可能与纹状体多巴胺的分泌减少有关，而不是源于简单的反应时-正确率权衡。总体而言，DDM 能从反应时和正确率中萃取出决策过程的不同成分，这是传统的行为数据分析方法所不能完成的。此外，根据观测数据拟合的 DDM 模型也可以用于估计驱动行为产生的不同成分的价值(Moustafa et al., 2015)。

与传统的群体平均数差异检验的方法相比，DDM 分析能提高结果的特异性(specificity)和敏感性(sensitivity)。在特异性方面，DDM 能鉴别出解释行为表现的决策成分。在敏感性方面，DDM 能发现行为结果中反应时和正确率的微小差异，甚至是不显著的差异(Pe et al., 2013; White et al., 2010a)。近期，研究者又发展出神经漂移扩散模型(neural drift diffusion model, NDDM)，它将行为数据与脑成像数据结合加以分析(Turner, van Maanen, & Forstmann, 2015)；层次漂移扩散模型(hierarchical drift diffusion model, HDDM)则通过层次贝叶斯方法(hierarchical Bayesian methods)，将被试和分组参数在不同水平上同时进行估计，能估计后验分布中参数的不确定性(Wiecki, Sofer, & Frank, 2013)。此外，Krajbich 和 Rangel(2011)将 DDM 推广至三重选择。更多 DDM 的原理、算法和工具包请见 van Ravenzwaaij 和 Oberauer(2009)、Ratcliff 和 McKoon(2008)、Wiecki 等人(2013)和 Ratcliff 等人(2016)的研究。

3.3 生成模型

贝叶斯大脑假设认为，大脑能对输入的感觉信息建构出一个生成模型，并不断进行更新，连续更新大脑的生成模型依赖于能量最小化(Stephan & Mathys, 2014)，即自由能量原则(free-energy principle)。自由能量原则指出，生物系统或个体在面对变化的环境时能维持其状态(Ashby, 1947)，尽量避免“惊奇”(“surprise”) 的产生所引发自由能量的波动，保证其心理状态在生理界限（自由能量）范围内(Friston, 2010)。具体而言，人的学习是由已有信念与新输入信息之间的不匹配引发的预测误差所驱动的，而精神疾病则是源于先前对感觉信息编码的精确性降低，基于不精确的感觉信息编码所形成的信念就产生了非适应性的推理(Sterzer et al., 2018)。

基于贝叶斯理论的生成模型，能够细致地测量不可观测的神经和认知过程(Huys, 2018b)。目前，最常用的生成模型是动态因果模型(dynamic causal modeling, DCM)。它关注于使用贝叶斯模型降阶(Bayesian model reduction)，即进行贝叶斯模型反演(Bayesian model inversion)和比较去巧妙处理单一数据集的多重模型或单一模型的多重数据集的反演(Friston et al., 2016)。基于生成模型和大脑功能连接假设，研究者将 DCM 应用于功能磁共振成像(functional magnetic resonance imaging, fMRI)和脑磁/脑电(magneto-/electroencephalography, M/EEG)数据的建模中。

DCM 首次被引入 fMRI 数据分析时，Friston 等人(2003)提出神经元活动的动态模型：

chinaXiv:201908.00074v1

$$\frac{dx}{dt} = \left(A + \sum_j B^{(j)} u_j\right) x + Cu$$

该双线性模型(bilinear model)描述的是，神经状态 x 的动态呈现为突触在神经节点或脑区（效应连接 A，endogenous connectivity A）和实验控制的操纵 u 对系统的作用。实验的操纵可能是直接影响神经元状态（直接输入 C），也可能是调节不同节点的效应连接（调节输入 B）(Stephan et al., 2008)。最佳 DCM 的选择由相关病理过程的特定假设决定。例如，非线性动态因果模型(nonlinear DCM)常用于区域性特定异常调节导致的异常突触可塑性问题(Frassle et al., 2018)，随机动态因果模型(stochastic DCM)和频谱动态因果模型(spectral DCM)则常用于检测静息态脑功能网络的异常(Friston et al., 2003; Friston, Kahan, Biswal, & Razi, 2014; Li et al., 2011; Razi, Kahan, Rees, & Friston, 2015)，而且频谱动态因果模型的预测精确性更高、对效应连接的组间差异检测更敏感(Almgren et al., 2018)。

已有不少研究者使用 DCM 对任务态 fMRI、静息态 fMRI、MEG 和 EEG 数据进行分析，并发现重性抑郁障碍患者在视觉、注意、情绪加工等相关的神经网络方面存在异常。在情绪加工方面，Lu 等人(2012)在重性抑郁障碍患者完成面孔情绪任务时记录其 MEG 信号。结果显示，重性抑郁障碍患者在背外侧前额叶至杏仁核自上而下的效应连接上存在明显减弱，而在杏仁核至前扣带回自下而上的效应连接，以及在前扣带回至背外侧前额叶的效应连接上存在明显增强。额叶对下级脑区的效应连接受损可能是导致重性抑郁障碍患者处理情绪信息功能受损的原因。在静息状态下，与控制组相比，重性抑郁障碍患者的右脑岛、右壳核和右尾状核(caudate)对喙部前扣带回的因果性效应连接增强，但双侧背外侧前额叶和左侧眶额皮层对喙部前扣带回的因果性效应连接降低，同样暗示着重性抑郁障碍患者存在认知控制的损伤(Feng et al., 2016)。在自我评价方面，重性抑郁障碍患者表现出内侧前额叶至后扣带回效应连接的过度激活，暗示着其自我加工对情绪的过度监管(“hyperregulation”)，这可能是重性抑郁障碍患者自我反思或反刍的重要原因(Davey, Breakspear, Pujol, & Harrison, 2018)。

总体而言，用 DCM 去拟合数据，将这些拟合的参数（如两个脑区的因果性连接强度）作为充分统计量(sufficient statistics)，能简洁地描述数据的特征，有望为抑郁症的诊断提供简洁有效的指标。从观测数据中生成的模型允许进行模型反演(model inversion)。既可以通过已有的心理现象来预测大脑活动，也允许根据大脑活动反推行为变化等。生成模型通过贝叶斯模型比较(Bayesian model comparison)等过程，对模型复杂性进行量化评估，增加了模型的稳健性、再生性和推广性(Huys, 2018a)。关于 DCM 的更多原理、基于 fMRI、MEG/EEG 的 DCM 及其在精神疾病的应用等请见 Frässle 等人(2017)、Friston 等人(2019)和 Stephan 等人(2010)的研究。

综合上述三种模型，通过计算模型对人的知觉、注意、执行控制和决策等过程进行模拟，可以将发现的神经生理基础和已有行为进行更细致的联结，模型中的参数（特定过程）也能为后续的研究和干预提供简洁有效的测量指标。而且，不同的模型有其独特的优势。例如，在行为层面，DDM 主要关注的是被试进行特定任务时的知觉、注意和简单的决策过程，强化学习模型则更关注奖赏和惩罚、价值加工对学习的动态影响；在神经层面，基于神经影像学数据的 DCM 能对静息态或任务态条件下被试不同脑区活动的因果关系进行模拟。因此，系统地研究抑郁症群体的认知功能需要不同模型的优势互补。虽然抑郁症的机制研究从未停止，数据在不断积累，方法也在不断更新，但关于其发病机制仍充满争议。因此，当存在大量未知影响的情况下，使用数据驱动的方法（如机器学习）来研究抑郁症可能是一种可行的方式。

chinaXiv:201908.00074v1

chinaXiv:201908.00074v1

4 数据驱动的计算精神病学研究

数据驱动的计算精神病学研究最常用的方法是机器学习。机器学习中最常用的类别是监督学习(supervised learning)和非监督学习(unsupervised learning)，主要包含分类、回归和聚类等三方面的方法。本文以监督学习方法为主，简要介绍这三类方法的原理，以及机器学习方法在抑郁症诊断、预测和治疗方案效果评价中的应用。

机器学习算法中，每一个数据集里的每一实例数据均被表征为同一组特征，这些特征可能是连续的、分类的或二元的。如果这些实例具有标签（已知的类别），那么该学习的过程被称为监督学习，否则该学习被称为非监督学习。监督学习包含了分类和回归的方法，基于分类的方法利用已有标签所区分的不同数据样本建立分类器，用于新样本的类别预测；基于回归的方法则将数据与一个连续函数进行拟合，再以该连续变量标记该数据。因此，监督学习常常通过构建分类器，来鉴别疾病群体和健康群体，或通过建立预测模型来预测抑郁症患病情况。

监督学习的主要步骤为特征提取（和选择）、模型训练和测试、模型评价。首先，获取被试（如患者 vs 健康组）的行为、基因、脑成像等数据，形成一个特征 X 为 N 名被试×P 个变量数据集，目标变量 Y 为已知的标签或类别，并进行特征提取(Bzdok & Meyer-Lindenberg, 2018)。研究者可以通过特征选择方法筛选感兴趣的特征，如使用 *t* 检验(t-test)方法进行分类，皮尔逊相关方法(Pearson correlation)进行回归等，也可以保留所有特征。然后，根据所选的算法进行模型训练和测试，如常用的支持向量机方法(support vector machine, SVM，见下文)。最后，进行模型评价，即通过训练样本得出的估计模型(in-sample estimate)来预测另一新样本的结果(out-of-sample estimate)。模型评价常用的方法是交叉验证法(cross-validation, CV)：将训练集分割成 K 个子样本，每个子样本将用于模型验证，其余的 $K-1$ 个样本则用于模型训练。每个子样本验证一次，对 K 次的模型结果进行平均或转换，最终得到一个单一的估计模型。因此，该方法通常又称为 K 次交叉验证(K-fold cross-validation)，其中 5 次或 10 次交叉验证方法的应用最为广泛(Bzdok & Meyer-Lindenberg, 2018)。基于分类的验证结果可以通过模型预测的正确率、敏感性、特异性、精确性、接受者操作特征曲线(receiver operating characteristic curve, ROC)以下区域的面积(area under ROC curve, AUC)等指标进行评价；基于回归的验证结果则需要通过相关系数、平均绝对误差、均方误差等进行评价(Bzdok & Meyer-Lindenberg, 2018)。

机器学习的具体算法众多，具体算法的选择由研究问题而定。目前，机器学习方法可分为分类算法、回归算法和聚类算法等三大类，其余分类方法和技术及其统计原理请见 Bzdok 和 Meyer-Lindenberg (2018)与 Shalev-Shwartz 和 Ben-David (2014)的介绍。在分类算法中，k-近邻算法、决策树算法、支持向量机和 AdaBoost 算法等均能对称名数据和数值型数据进行分类。其中，支持向量机方法应用较为广泛。支持向量机就是一个有效区分出训练集不同类别样本的超平面（n 维平面），并能保证该超平面和最靠近超平面的几个训练样本点之间间隔最大的工具(Andrews, Tsochantaridis, & Hofmann, 2002)。该算法具有较高的分类准确性和较好的泛化能力，能避免出现数据过度拟合的现象(Singh, Thakur, & Sharma, 2016)。

在回归算法中，线性回归算法应用较多，主要包括最小绝对值收敛和选择算法(least absolute shrinkage and selection operator, LASSO)，岭回归(ridge regression)，相关向量回归(relevance vector regression, RVR)和弹性网回归(elastic net regression)等。Cui 和 Gong(2018)对不同回归算法的计算耗时和预测准确性进行了评价，发现岭回归时效性和准确性较高。岭回归模型使预测误差和回归系数的平方最小化，通过缩减(shrinkage)来去掉不重要的参数，能较好地处理多重共线性和过度拟合问题(Cui & Gong, 2018)。

分类和回归方法常用于监督学习，而聚类法则常用于非监督学习中，其主要目的是寻

找数据集的内在分布结构。k-均值算法是最常见的聚类算法之一。它需要用户给定的 k 个随机质心，每一个样本点会被分配到与它最近的簇质心，引起簇质心的改变。该过程不断重复，直至簇质心不再变化(Wagstaff, Cardie, Rogers, & Schrödl, 2001)。此外，更多更高效的聚类算法，如二分 k-均值算法和层次聚类算法等将不在此一一介绍。

目前，基于上述算法形成的监督学习方法和非监督学习方法已经广泛地应用于抑郁症的诊断、预测和治疗方案的评价。下文将主要介绍监督学习方法在临床中的应用。

4.1 疾病诊断和预测

监督学习应用于精神疾病的诊断和预测，主要包括三个阶段。首先，研究者编写出其认为能成功学习的算法；其次，使用该算法对（行为、生理、脑成像等）数据进行训练，计算学习结果与真实结果（是否患精神病）的差异或距离，通过调节内部参数（权重，weight）来尽可能地缩小该差异，产生最优的学习结果，并将该数据特征标记为某精神疾病；最后，使用该模型对一些新数据（可能是患者或非患者的数据）进行预测(LeCun, Bengio, & Hinton, 2015)。如果学习成功，那么所有或大多数标记将被正确识别。此外，除了上述主要步骤以外，研究者通常还会对结果进行模式定位(pattern localization)，例如对特定的或重要的向量进行解释，或根据不同特征对标签预测的权重绘制出可视化的图形——权重地图(weight map)。

近年来，已有不少研究通过机器学习方法对重性抑郁障碍进行诊断和预测。例如，Patel 等人(2015)通过监督学习方法，对被试的年龄、简易精神状态检查得分和结构磁共振成像数据，使用交替决策树算法(alternating decision tree)预测老年抑郁(late-life depression)，敏感性指数为 88.89%，特异性指数为 85.71%，预测准确率较高。有研究利用微博等社交媒体中用户的文字表达来预测用户是否患有抑郁症。与朴素贝叶斯算法和决策树算法等相比，支持向量机算法的预测敏感性和特异性均较高，分别为 83.3%和 82.6% (Haque, Guo, Miner, & Li, 2018)。另有研究使用多视图双聚类算法，以多视角感知数据为输入信息，进而建立模型来鉴别抑郁组群体和健康组群体，并使用支持向量机分类器对该模型进行验证，结果显示正确率达 87% (Farhan et al., 2016)。此外，还有研究者通过非监督学习方法对抑郁症进行诊断和预测。Drysdale 等人(2017)对 1188 名抑郁症患者进行了静息态功能磁共振成像，并对患者在边缘系统网络和额叶-纹状体网络的异常功能连接数据进行了层次聚类分析(hierarchical clustering)，产生了四种神经生理类型（诊断分类器）。然后，经过多位点验证(multisite validation)和新样本验证(out-of-sample validation)，发现聚类具有较高的敏感性和特异性(82~93%)。

然而，上述研究均是横断研究，检验预测准确度更好的方法应是通过纵向研究，利用生成的机器学习模型，基于个体的过往数据来预测其未来罹患抑郁症的风险，并与现实情况进行对比，验证该模型的预测准确度。此外，近年来有越来越多研究关注于单个患者的脑成像数据如何预测其未来患精神疾病的概率，这涉及贝叶斯模型选择(Bayesian model selection)和生成嵌入(generative embedding)等过程，具体详见 Stephan 等人(2017)的研究。

4.2 治疗效果预测与方案选择

研究者并不满足于当前机器学习在鉴别抑郁症患者和健康人之中的作用，他们还将该方法推广至治疗效果的判定与治疗方案的选择。过往研究表明，不同抗抑郁药在不同患者中疗效不一，但这种个体差异对不同患者选择个性化的治疗方案十分重要。目前已有研究开始使用单变量或多变量标记物来预测治疗效果。例如，DeRubeis 等人(2014)的研究检测了不同疗法对不同患者群体的疗效，发现已婚、就业、具有丰富阅历但对抗抑郁药物产生耐受性的个体能通过认知行为疗法(cognitive behavioral therapy, CBT)缓解抑郁症状，而人格障碍和抑郁障碍的共病个体服用抗抑郁药缓解抑郁症状更有效。Williams 等人(2015)利用单一的标记物，即杏仁核对阈上和阈下情绪面孔刺激的反应，去预测特定药物对不同抑郁症

chinaXiv:201908.00074v1

患者的疗效。结果显示，患者的杏仁核在加工阈下奖赏性和威胁性面孔时激活减弱，这一信号表明该患者对抗抑郁药存在普遍的敏感性(Cohen’s $d = 0.63 \sim 0.77$)；患者的杏仁核在加工阈下悲伤面孔时过度激活，则表明该患者对 5-羟色胺–去甲肾上腺素重摄取抑制剂产生耐药性(Cohen’s $d = 1.5$)。因此，阈下面孔加工这一认知标记物能有效预测患者对特定抗抑郁药的敏感性。

除了利用单一标记物检验和预测药物的治疗效果，目前有较多研究尝试使用更多维的标记物去提高预测的准确性。Rush 等人(2006)所实施的抑郁症疗效对比的纵向研究指出，在约 4000 名重性抑郁障碍患者中，50%以上患者对西酞普兰药物敏感，在后续阶段其缓解率较高、复发率较低，而不敏感者即使更换其他药物，其后续阶段的症状缓解率仍然较低、复发率较高。另外，有研究者分析了 4041 名抑郁症患者的症状，使用机器学习模型去判定特定抗抑郁药对哪些个体疗效最佳。最后生成的模型同时考虑了 164 个潜变量，涉及躯体症状、失眠、过往创伤事件的暴露时间等多方面，得到了较好的预测准确度(Chekroud, Gueorguieva, et al., 2017; Chekroud et al., 2016)。

基于治疗效果的预测，使用机器学习方法可以为患者建立最佳的治疗方案。DeBattista 等人(2011)通过使用参考脑电图(referenced-electroencephalogram, rEEG)来预测药物对患者的疗效。与 Rush 等人(2006)基于药物疗效更换药物的治疗方案相比，使用基于 rEEG 数据（超过 1800 名患者、总长达 405 天、超过 17000 个药物试次、包含 74 个可用生物标记物的样本），通过自动治疗选择算法(automatic treatment-selection algorithms)选择合适药物的方法所得的治疗效果显著提升。若该结果能得到更多验证，基于 rEEG 的方案将会大大提高治疗的效率，并成为简易廉价且兼具客观性和预测性的抗抑郁药物选择程序。

总体而言，呈指数型增长的基于人工智能的精神疾病研究已经在抑郁症的诊断、预测和治疗等方面表现出一定的成效(Tran et al., 2019)。目前，抑郁症的诊断和治疗通常利用神经影像学数据、可穿戴设备的传感器数据、社交媒体数据和调查数据，以支持向量机方法和回归算法来建立分类器或预测模型(Shatte, Hutchinson, & Teague, 2019)，这些模型预测准确率较高。另外，在抑郁症治疗方案的评价方面，不少研究者以神经成像数据、临床治疗数据和调查数据，以及常见的药物和生物标记物的数据，通过纵向追踪的方式来检测不同算法训练得出的预测模型的准确性。目前，这类方法在临床中的应用不断增加，再进一步的推广则需更成熟的算法和更深入的抑郁症病理机制研究。

chinaXiv:201908.00074v1

5 计算精神病学的优势与局限

5.1 优势

5.1.1 关注于个体

以往研究通常将抑郁组和健康组被试进行对比，以发现抑郁组被试异常的生理心理功能，目前已有一定的研究成果（如孙也婷等，印刷中；文宏伟，陆菁菁，何晖光，2018）。然而，随着研究的深入和新技术的使用，这种基于群体比较的方法对于研究复杂的精神疾病而言效率较低，结果也不尽如人意。抑郁症的产生源于基因、认知和情绪、环境等因素的交互影响，每种因素的效应因个体的情况（如年龄、个体经验、家庭环境，甚至文化背景）而异。因此，抑郁症的治疗方案的效果也可能因人而异。所以，针对成年早期抑郁症患者的治疗方案不一定适用于老年人和青少年(Chekroud, Lane, & Ross, 2017)。与传统的精神病学研究相比，计算精神病学的方法更关注于个体，能鉴别出具有抑郁风险的个体，并可以根据其预测模型定位至可能导致抑郁的某些核心特征。因此，这种“个性化治疗”能更高效地帮助个体在生理和行为上得到改善(Chekroud, Lane, et al., 2017)。

5.1.2 充分利用研究数据

Robinson 和 Chase(2017)认为，使用计算建模方法的前提是，研究假设是定量描述的，

这有利于用参数去模拟单一或多个分离的过程。通过该方法研究人类行为能更充分地利用所获得的数据，并能更准确地预测人类行为。首先，多个预测模型之间的相互比较能区分出模型是否适用并能估算出这些模型的拟合强度(Robinson & Chase, 2017)。其次，对于精神疾病患者而言，认知和学习过程的研究尤为重要，该过程与不少精神疾病的出现和维持紧密相关，包括抑郁症(Rock, Roiser, Riedel, & Blackwell, 2014)、焦虑症(Gilmartin, Balderston, & Helmstetter, 2014)、精神分裂症(Cicero, Martin, Becker, & Kerns, 2014)等。学习的过程是不断变化的，需要监控每个试次的选择及其变异(trial-by-trial variance)，计算建模恰恰能做到这一点。而传统心理学的分析方法一般是通过反应时和正确率的平均数和标准差来反映数据的特征，忽略了一些微小的或动态的变化。再次，计算模型能将行为观测指标和生理、神经层面的变化联系在一起，为解释行为产生的内部机制提供有价值的信息。最后，计算模型为复杂的认知和决策过程提供了一些参数，如病理学中常用的元参数(meta parameters)或模型参数的变异(variance of model parameters)。

综上所述，计算建模方法对多层面、多系统的数据进行处理，而非基于单一的自我报告和临床访谈，促进了研究者对抑郁症病理机制的认识，提高了诊疗的客观性。其次，计算建模方法能针对个体的抑郁程度和症状特征等（如共病）进行细致地描述，一定程度上克服了由共病性和异质性引起的诊疗的困难。最后，计算模型的参数可以为特定的心理过程或症状表现（如奖赏加工的学习率）提供定量的描述，而非二分的结果，临床医师可以根据该数值来判断个体的抑郁程度，并提出更精准的治疗方案。

5.2 不足与展望

5.2.1 理论驱动的研究

尽管以计算建模方法研究抑郁症等精神疾病的项目已逐渐增多，并且该方法能从生物分子到行为环境等层面来解释不同水平及其交互作用如何影响了抑郁症的产生、发展、维持和康复，但其有效性仍受到不少人的质疑——计算建模是否足够成熟，并足以对精神疾病进行良好的判定和预测(Huys, Moutoussis, & Williams, 2011)？

首先，模型确实能从参数的不同赋值中揭示出患者的反应偏差、偏好或价值更新过程(Wiecki, Poland, & Frank, 2015)，但这些参数并不能涵盖某一心理过程的所有方面。其次，模型中的某些参数的生物学意义模糊。不同的研究者会从不同的角度来理解同一个参数，但这种参数究竟能否匹配精神病学、心理学或神经生物学的概念(Huys et al., 2011)？例如，在强化学习模型中，Husain 和 Roiser(2018)提出，基于强化学习模型的 Softmax 函数中，β 值可以被解释为不一致性或探索性，但究竟 β 值代表的是哪一个心理过程？又对应着哪些神经生理指标？再次，理论驱动的研究所发现的结果是否具有可重复性(Maia, Huys, & Frank, 2017)？如上文所述，无模型学习过程中，抑郁症群体加工奖赏预测误差的功能是正常的还是异常的？上述问题的解决既依赖于数学模型本身的发展，如模型参数反映的生理心理过程是否明确，考察数据的维度是否全面等，也依赖于不同学科的研究者的通力合作，从基因、细胞等分子层面、神经和内分泌系统层面，以及行为和外界环境层面，更新抑郁症动态发展的病理机制模型。最后，以往研究使用生理、脑成像与行为指标等描述了抑郁症患者认知、决策等的特点，但这些研究往往揭示的只是相关关系(Etkin, 2018)。在临床神经科学上，使用 TMS 或经颅直流电刺激(transcranial direct-current stimulation, tDCS)、进行药物操纵或治疗，结合成像技术，可以揭示生物分子、神经系统和行为等多层面、多系统之间的因果关系，如 rTMS 和 tDCS 刺激左侧背外侧前额叶也能缓解抑郁或双相障碍的症状(Boes et al., 2018; Donde et al., 2017)等。

5.2.2 数据驱动的研究

采用传统研究方法的抑郁症研究进展较慢，这可能是由于抑郁症等疾病存在较大的复杂性和异质性，而机器学习方法能从复杂的现象中提取主要特征，这也许能在一定程度上

chinaXiv:201908.00074v1

克服抑郁症的复杂性和异质性所带来的困难。目前，通过机器学习方法对抑郁症等精神疾病的诊断和疗效评估研究已初见成效，但仍存在一些局限与不足。首先，在监督学习方法中，样本的标签对机器学习的建模十分重要。例如，研究者通常是使用临床量表的方式对个体是否患有抑郁症进行判断，如汉密顿抑郁量表(Janssen, Mourao-Miranda, & Schnack, 2018)，而上文已提到，基于症状学的疾病诊断效果不佳，使用该结果判定样本的标签就已经存在了症状学方法的缺陷，该标签可能本身就是不准确的，采用不准确的标签进而希望得到准确的预测结果，可能不太可行。而且，单一的分类或连续性标签可能会丢失很多额外的信息（如是否存在共病情况等），不能较为全面地反映出该类别数据的特征，容易使机器学习过程产生偏差，从而导致模型预测的结果与现实结果存在差距。鉴于此，有研究者提出采用非监督学习方法以改善这一现状。非监督学习基于已有的数据集，不需要标签的输入，通过聚类、关联分析等方法对数据进行降维，减少了人为定义或分类所引起的偏差。此外，在样本量很大和数据维度较多时，深度学习方法（如递归神经网络、卷积神经网络等）通过自己学习样本特征，不需要人为编码，并且由学习形成的网状层级结构，能在一定程度上提高分类和预测的准确性。因此，对于复杂而异质的精神疾病（如抑郁症）的鉴别而言，精准的标签、充分的信息量和合适的方法能更好地解释出现疾病的各种亚型和共病现象的原因，进而制定相应的诊疗方案和预防措施，这可能也是未来发展的趋势之一。

其次，机器学习需要大样本，样本量过小较易使模型识别出现较大的测量误差和模型变异性(Schnack & Kahn, 2016)，而研究数据的共享能较好地解决样本量问题；另外，样本过于同质会导致模型识别得到的结果与样本相关而非与疾病本身相关(Mendelson et al., 2017)。这就引出了新的问题，即采用不同的成像仪器或应用于新样本时，原有模型的预测效力能否保持？由于机器学习对模型的交叉验证使用的是两个同源的子样本，因此，在不同来源的新样本中，如不同性别、年龄、教育程度和文化背景的样本等，所产生的模型是否具有推广性和可靠性，这仍需要采集全新的样本来加以验证。此外，研究者为了文章的发表，可能只选择性地报告了预测效力高的模型，机器学习模型的预测正确率存在被高估的风险，这也需引起研究者的重视。

5.2.3 结合理论驱动和数据驱动的方法

数据驱动和理论驱动方法都是根据一个测量获得的数据集，进行合适的统计和推断的过程，进而得出结论。理论驱动的方法只是在选定测量指标时依据研究者的经验和知识进行了数据的降维；而数据驱动的方法则通常尽可能地测量所有可测量的变量，将这些指标利用机器学习方法进行降维或全局分析，其结果的解释率或预测率相对较高。目前，研究者对抑郁症等精神疾病的理论认识尚浅，过多依赖于过往经验往往会引发结果的偏差，导致解释率偏低。然而，根据机器学习所得到的维度或因子通常不易被解释，测量过程需要较大的工作量，这是该方法不能较好应用于研究和临床领域的重要原因。Huys 等人(2016)认为，当理论驱动的模型参数是充分统计量时，它能捕获驱动复杂观测过程的基本模式，使数据驱动的功效最大化。因此，结合理论驱动方法和数据驱动方法可以提高研究与应用的效率和可靠性(Huys et al., 2016)。研究者可以根据机器学习算法得出的降维结果，整合已有知识经验，分析维度产生的原因及各维度的贡献率，深入认识抑郁症的发病机制及核心特征，为抑郁症的防治提供理论指导。研究者还可以基于更成熟、更合适的算法，如深度学习、（广泛意义上的）强化学习，甚至迁移学习，对抑郁症的病理机制进行更好地拟合或建构，为抑郁症的预防和诊疗提供精准的方案。

5.2.4 人才和资源的整合

计算精神病学研究需要跨学科、多方法融合，但很少有人能精通多门学科，因此这类研究迫切需要跨学科人才的合作与交流，包括实验室的共建、仪器和技术的共享、研究方

chinaXiv:201908.00074v1

法的互补等（谢小华，冯建峰，2019）。

抑郁症的计算精神病学研究致力于生成新的诊断图式、提出精准的治疗方案。但这一目标的实现仍需要时间(Chekroud, Lane, et al., 2017)。需要强调的是，计算精神病学的发展离不开药物的发展，离不开生物标记物、药物和医学设备的发展循环，离不开前临床研究、临床化验发展、纵向和横向的研究、前瞻性研究和随机对照试验(Bahn, Noll, Barnes, Schwarz, & Guest, 2011; Czajkowski et al., 2015; Drucker & Krapfenbauer, 2013)等的发展与互补。计算精神病学的研究者需培养跨学科的意识，了解不同学科的方法和最新进展，发挥本学科的特长，与其他学科的研究者通力合作，从多层面深入认识抑郁症的发病机制。

另外，数据和代码资源的共享也十分重要。数据驱动研究的数据集维度较大，在统计上对样本量的需求也相应变大。单个研究者的资源有限，所获得的数据量往往达不到要求。如果想利用机器学习获得准确的预测模型，研究者在增大样本量的同时还需收集多个层面的信息。目前，从基因、神经递质、神经系统、大脑等至行为、环境等多个层面收集患者的数据需要花费较大的人力物力，因此，迫切需要数据资源的共享。这就要求多个领域的研究者按照相同的标准和参数收集患者的生理、脑成像、行为指标等数据，附上数据分析的代码，共享给其他研究者。这种方式不仅能提高数据资源的利用率，还能推动各类算法的优化，并能促进计算精神病学研究结果的可重复性检验。

6 结论

综上所述，计算精神病学研究促进了研究者对抑郁症发病机制和患者的认知加工过程的深入认识。计算精神病学方法关注于个体、充分利用个体的多层面数据，使得诊断具有客观、精准和个体特异的优势，但也存在模型参数解释困难、算法优化不足和数据采集不便等缺点。然而，只要不同学科的研究者发挥其专长，合作交流、资源共享、方法革新、优势互补，抑郁症的计算精神病学研究有望能从微观至宏观等多层面多系统之间的交互作用中揭示抑郁症的病理，其研究成果有望转化为对抑郁症患者的精准防治，最终降低抑郁症的发病率，提升人们的心理健康水平。

参考文献

孙也婷, 陈桃林, 何度, 董再全, 程勃超, 王淞, . . . 龚启勇. (印刷中) 基于精神影像和人工智能的抑郁症客观生物标志物研究进展. *生物化学与生物物理进展*, 1-45.

文宏伟, 陆菁菁, 何晖光. (2018). 机器学习在神经精神疾病诊断及预测中的应用. *协和医学杂志, 9*(1), 19-24.

谢小华, 冯建峰. (2019). 上海市脑与类脑智能基础转化应用研究的现状及展望. *心理学通讯, 2*(02), 84-87.

Almgren, H., Van de Steen, F., Kuhn, S., Razi, A., Friston, K., & Marinazzo, D. (2018). Variability and reliability of effective connectivity within the core default mode network: A multi-site longitudinal spectral DCM study. *Neuroimage, 183*, 757–768. doi:10.1016/j.neuroimage.2018.08.053

Andrews, S., Tsochantaridis, I., & Hofmann, T. (2002). *Support vector machines for multiple-instance learning*. Paper presented at the Advances in Neural Information Processing Systems 15, Vancouver, British Columbia, Canada.

Arnsten, A. F. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience, 10*(6), 410–422. doi:10.1038/nrn2648

Ashby, W. R. (1947). Principles of the self-organizing dynamic system. *Journal of General Psychology, 37*(2), 125–128. doi:10.1080/00221309.1947.9918144

chinaXiv:201908.00074v1

Association, A. P. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*: American Psychiatric Pub.

Bahn, S., Noll, R., Barnes, A., Schwarz, E., & Guest, P. C. (2011). Challenges of introducing new biomarker products for neuropsychiatric disorders into the market. *International Review of Neurobiology, 101*, 299–327. doi:10.1016/B978-0-12-387718-5.00012-2

Boes, A. D., Uitermarkt, B. D., Albazron, F. M., Lan, M. J., Liston, C., Pascual-Leone, A., . . . Fox, M. D. (2018). Rostral anterior cingulate cortex is a structural correlate of repetitive TMS treatment response in depression. *Brain Stimulation, 11*(3), 575–581. doi:10.1016/j.brs.2018.01.029

Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(3), 223–230. doi:10.1016/j.bpsc.2017.11.007

Chekroud, A. M. (2017). Bigger Data, Harder Questions-Opportunities Throughout Mental Health Care. *JAMA Psychiatry, 74*(12), 1183–1184. doi:10.1001/jamapsychiatry.2017.3333

Chekroud, A. M., Gueorguieva, R., Krumholz, H. M., Trivedi, M. H., Krystal, J. H., & McCarthy, G. (2017). Reevaluating the Efficacy and Predictability of Antidepressant Treatments: A Symptom Clustering Approach. *JAMA Psychiatry, 74*(4), 370–378. doi:10.1001/jamapsychiatry.2017.0025

Chekroud, A. M., Lane, C. E., & Ross, D. A. (2017). Computational Psychiatry: Embracing Uncertainty and Focusing on Individuals, Not Averages. *Biological Psychiatry, 82*(6), e45–e47. doi:10.1016/j.biopsych.2017.07.011

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., . . . Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: A machine learning approach. *Lancet Psychiatry, 3*(3), 243–250. doi:10.1016/S2215-0366(15)00471-X

Chen, C., Takahashi, T., Nakagawa, S., Inoue, T., & Kusumi, I. (2015). Reinforcement learning in depression: A review of computational research. *Neuroscience and Biobehavioral Reviews, 55*, 247–267. doi:10.1016/j.neubiorev.2015.05.005

Cicero, D. C., Martin, E. A., Becker, T. M., & Kerns, J. G. (2014). Reinforcement learning deficits in people with schizophrenia persist after extended trials. *Psychiatry Research, 220*(3), 760–764. doi:10.1016/j.psychres.2014.08.013

Clark, L., Chamberlain, S. R., & Sahakian, B. J. (2009). Neurocognitive mechanisms in depression: Implications for treatment. *Annual Review of Neuroscience, 32*(1), 57–74. doi:10.1146/annurev.neuro.31.060407.125618

Cooper, J. A., Arulpragasam, A. R., & Treadway, M. T. (2018). Anhedonia in depression: biological mechanisms and computational models. *Current Opinion in Behavioral Sciences, 22*, 128–135. doi:10.1016/j.cobeha.2018.01.024

Cui, Z., & Gong, G. (2018). The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *Neuroimage, 178*, 622–637. doi:10.1016/j.neuroimage.2018.06.001

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine, 11*(1), 126. doi:10.1186/1741-7015-11-126

Czajkowski, S. M., Powell, L. H., Adler, N., Naar-King, S., Reynolds, K. D., Hunter, C. M., . . . Charlson, M. E. (2015). From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. *Health Psychology, 34*(10), 971–982.

chinaXiv:201908.00074v1

doi:10.1037/hea0000161

Davey, C., Breakspear, M., Pujol, J., & Harrison, B. (2018). 201. A Dynamic Causal Model of the Depressed Self. *Biological Psychiatry, 83*(9). doi:10.1016/j.biopsych.2018.02.220

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69*(6), 1204–1215. doi:10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*(12), 1704–1711. doi:10.1038/nn1560

DeBattista, C., Kinrys, G., Hoffman, D., Goldstein, C., Zajecka, J., Kocsis, J., . . . Fava, M. (2011). The use of referenced-EEG (rEEG) in assisting medication selection for the treatment of depression. *Journal of Psychiatric Research, 45*(1), 64–75. doi:10.1016/j.jpsychires.2010.05.009

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The Personalized Advantage Index: Translating research on prediction into individualized treatment recommendations. A demonstration. *PloS One, 9*(1), e83875. doi:10.1371/journal.pone.0083875

Dillon, D. G., Wiecki, T., Pechtel, P., Webb, C., Goer, F., Murray, L., . . . Pizzagalli, D. A. (2015). A computational analysis of flanker interference in depression. *Psychological Medicine, 45*(11), 2333–2344. doi:10.1017/S0033291715000276

Doll, B. B., Bath, K. G., Daw, N. D., & Frank, M. J. (2016). Variability in Dopamine Genes Dissociates Model-Based and Model-Free Reinforcement Learning. *Journal of Neuroscience, 36*(4), 1211–1222. doi:10.1523/JNEUROSCI.1901-15.2016

Donde, C., Amad, A., Nieto, I., Brunoni, A. R., Neufeld, N. H., Bellivier, F., . . . Geoffroy, P. A. (2017). Transcranial direct-current stimulation (tDCS) for bipolar depression: A systematic review and meta-analysis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 78*, 123–131. doi:10.1016/j.pnpbp.2017.05.021

Drucker, E., & Krapfenbauer, K. (2013). Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. *The EPMA Journal 4*(1), 7. doi:10.1186/1878-5085-4-7

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., . . . Liston, C. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine, 23*(1), 28–38. doi:10.1038/nm.4246

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E. J. (2012). Testing theories of post-error slowing. *Attention Perception & Psychophysics, 74*(2), 454–465. doi:10.3758/s13414-011-0243-2

Eldar, E., Roth, C., Dayan, P., & Dolan, R. J. (2018). Decodability of Reward Learning Signals Predicts Mood Fluctuations. *Current Biology, 28*(9), 1433–1439 e1437. doi:10.1016/j.cub.2018.03.038

Etkin, A. (2018). Addressing the Causality Gap in Human Psychiatric Neuroscience. *JAMA Psychiatry, 75*(1), 3–4. doi:10.1001/jamapsychiatry.2017.3610

Farhan, A. A., Lu, J., Bi, J., Russell, A., Wang, B., & Bamis, A. (2016). *Multi-view Bi-clustering to Identify Smartphone Sensing Features Indicative of Depression.* Paper presented at the 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Washington, DC, USA.

chinaXiv:201908.00074v1

Feng, Z., Xu, S., Huang, M., Shi, Y., Xiong, B., & Yang, H. (2016). Disrupted causal connectivity anchored on the anterior cingulate cortex in first-episode medication-naive major depressive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 64*, 124–130. doi:10.1016/j.pnpbp.2015.07.008

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annual Review of Psychology, 67*, 641–666. doi:10.1146/annurev-psych-122414-033645

Frassle, S., Lomakina, E. I., Razi, A., Friston, K. J., Buhmann, J. M., & Stephan, K. E. (2017). Regression DCM for fMRI. *Neuroimage, 155*, 406–421. doi:10.1016/j.neuroimage.2017.02.090

Frassle, S., Yao, Y., Schobi, D., Aponte, E. A., Heinzle, J., & Stephan, K. E. (2018). Generative models for clinical applications in computational psychiatry. *Wiley Interdisciplinary Reviews: Cognitive Science, 9*(3), e1460. doi:10.1002/wcs.1460

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138. doi:10.1038/nrn2787

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage, 19*(4), 1273–1302. doi:10.1016/s1053-8119(03)00202-7

Friston, K. J., Kahan, J., Biswal, B., & Razi, A. (2014). A DCM for resting state fMRI. *Neuroimage, 94*, 396–407. doi:10.1016/j.neuroimage.2013.12.009

Friston, K. J., Litvak, V., Oswal, A., Razi, A., Stephan, K. E., van Wijk, B. C. M., . . . Zeidman, P. (2016). Bayesian model reduction and empirical Bayes for group (DCM) studies. *Neuroimage, 128*, 413–431. doi:10.1016/j.neuroimage.2015.11.015

Friston, K. J., Preller, K. H., Mathys, C., Cagnan, H., Heinzle, J., Razi, A., & Zeidman, P. (2019). Dynamic causal modelling revisited. *Neuroimage, 199*, 730–744. doi:10.1016/j.neuroimage.2017.02.045

Gilmartin, M. R., Balderston, N. L., & Helmstetter, F. J. (2014). Prefrontal cortical regulation of fear learning. *Trends in Neurosciences, 37*(8), 455–464. doi:10.1016/j.tins.2014.05.004

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*(1), 535–574. doi:10.1146/annurev.neuro.29.051605.113038

Goldberg, D., & Fawcett, J. (2012). The importance of anxiety in both major depression and bipolar disorder. *Depression and Anxiety, 29*(6), 471–478. doi:10.1002/da.21939

Gomez, P., & Perea, M. (2014). Decomposing encoding and decisional components in visual-word recognition: A diffusion model analysis. *Quarterly Journal of Experimental Psychology (2006), 67*(12), 2455–2466. doi:10.1080/17470218.2014.937447

Hammen, C. (2018). Risk Factors for Depression: An Autobiographical Review. *Annual Review of Clinical Psychology, 14*, 1–28. doi:10.1146/annurev-clinpsy-050817-084811

Hanks, T. D., Ditterich, J., & Shadlen, M. N. (2006). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nature Neuroscience, 9*(5), 682–689. doi:10.1038/nn1683

Haque, A., Guo, M., Miner, A. S., & Li, F.-F. (2018). Measuring Depression Symptom Severity from Spoken Language and 3D Facial Expressions. *arXiv preprint arXiv:1811.08592*.

Heller, A. S., Ezie, C. E. C., Otto, A. R., & Timpano, K. R. (2018). Model-based learning and individual differences in depression: The moderating role of stress. *Behaviour Research and Therapy, 111*, 19–26. doi:10.1016/j.brat.2018.09.007

chinaXiv:201908.00074v1

Herrman, H., Kieling, C., McGorry, P., Horton, R., Sargent, J., & Patel, V. (2019). Reducing the global burden of depression: A Lancet–World Psychiatric Association Commission. *The Lancet, 393*(10189), e42–e43. doi:10.1016/s0140-6736(18)32408-5

Honnorat, N., Dong, A., Meisenzahl-Lechner, E., Koutsouleris, N., & Davatzikos, C. (2017). Neuroanatomical heterogeneity of schizophrenia revealed by semi-supervised machine learning methods. *Schizophrenia Research*. doi:10.1016/j.schres.2017.12.008

Husain, M., & Roiser, J. P. (2018). Neuroscience of apathy and anhedonia: A transdiagnostic approach. *Nature Reviews Neuroscience, 19*(8), 470–484. doi:10.1038/s41583-018-0029-9

Huys, Q. J. M. (2014). Computational Psychiatry. In *Encyclopedia of Computational Neuroscience* (pp. 1–10).

Huys, Q. J. M. (2018a). Advancing Clinical Improvements for Patients Using the Theory-Driven and Data-Driven Branches of Computational Psychiatry. *JAMA Psychiatry, 75*(3), 225–226. doi:10.1001/jamapsychiatry.2017.4246

Huys, Q. J. M. (2018b). Bayesian Approaches to Learning and Decision-Making. In A. Anticevic & J. D. Murray (Eds.), *Computational Psychiatry* (pp. 247–271): Academic Press.

Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience, 19*(3), 404–413. doi:10.1038/nn.4238

Huys, Q. J. M., Moutoussis, M., & Williams, J. (2011). Are computational models of any use to psychiatry? *Neural Networks, 24*(6), 544–551. doi:10.1016/j.neunet.2011.03.001

Insel, T. R., & Cuthbert, B. N. (2015). Medicine. Brain disorders? Precisely. *Science, 348*(6234), 499–500. doi:10.1126/science.aab2358

Janssen, R. J., Mourao-Miranda, J., & Schnack, H. G. (2018). Making Individual Prognoses in Psychiatry Using Neuroimaging and Machine Learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(9), 798–808. doi:10.1016/j.bpsc.2018.04.004

Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry, 17*(12), 1174–1179. doi:10.1038/mp.2012.105

Kendler, K. S. (2008). Explanatory models for psychiatric illness. *American Journal of Psychiatry, 165*(6), 695–702. doi:10.1176/appi.ajp.2008.07071061

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America, 108*(33), 13852–13857. doi:10.1073/pnas.1101328108

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. doi:10.1038/nature14539

Ledford, H. (2014). Medical research: If depression were cancer. *Nature, 515*(7526), 182–184. doi:10.1038/515182a

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual Review of Neuroscience, 35*(1), 287–308. doi:10.1146/annurev-neuro-062111-150512

Li, B., Daunizeau, J., Stephan, K. E., Penny, W., Hu, D., & Friston, K. (2011). Generalised filtering and stochastic DCM for fMRI. *Neuroimage, 58*(2), 442–457. doi:10.1016/j.neuroimage.2011.01.085

Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics.

chinaXiv:201908.00074v1

chinaXiv:201908.00074v1

Nature Reviews Genetics, 16(6), 321–332. doi:10.1038/nrg3920

Lin, T., Liu, T., Lin, Y., Yan, L., Chen, Z., & Wang, J. (2017). Comparative study on serum levels of macro and trace elements in schizophrenia based on supervised learning methods. *Journal of Trace Elements in Medicine and Biology, 43*, 202–208. doi:10.1016/j.jtemb.2017.03.010

Lu, Q., Li, H., Luo, G., Wang, Y., Tang, H., Han, L., & Yao, Z. (2012). Impaired prefrontal-amygdala effective connectivity is responsible for the dysfunction of emotion process in major depressive disorder: a dynamic causal modeling study on MEG. *Neuroscience Letters, 523*(2), 125–130. doi:10.1016/j.neulet.2012.06.058

Lu, Y., Tang, C., Liow, C. S., Ng, W. W., Ho, C. S., & Ho, R. C. (2014). A regressional analysis of maladaptive rumination, illness perception and negative emotional outcomes in Asian patients suffering from depressive disorder. *Asian Journal of Psychiatry, 12*, 69–76. doi:10.1016/j.ajp.2014.06.014

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience, 14*(2), 154–162. doi:10.1038/nn.2723

Maia, T. V., Huys, Q. J. M., & Frank, M. J. (2017). Theory-Based Computational Psychiatry. *Biological Psychiatry, 82*(6), 382–384. doi:10.1016/j.biopsych.2017.07.016

Malhi, G. S., & Mann, J. J. (2018). Depression. *The Lancet, 392*(10161), 2299–2312. doi:10.1016/s0140-6736(18)31948-2

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex, 13*(11), 1257–1269. doi:10.1093/cercor/bhg097

McEwen, B. S., & Morrison, J. H. (2013). The brain on stress: Vulnerability and plasticity of the prefrontal cortex over the life course. *Neuron, 79*(1), 16–29. doi:10.1016/j.neuron.2013.06.028

Mendelson, A. F., Zuluaga, M. A., Lorenzi, M., Hutton, B. F., Ourselin, S., & Alzheimer's Disease Neuroimaging, I. (2017). Selection bias in the reported performances of AD classification pipelines. *Neuroimage: Clinical, 14*, 400–416. doi:10.1016/j.nicl.2016.12.018

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences, 16*(1), 72–80. doi:10.1016/j.tics.2011.11.018

Moustafa, A. A., Keri, S., Somlai, Z., Balsdon, T., Frydecka, D., Misiak, B., & White, C. (2015). Drift diffusion model of reward and punishment learning in schizophrenia: Modeling and experimental data. *Behavioural Brain Research, 291*, 147–154. doi:10.1016/j.bbr.2015.05.024

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology, 53*(3), 139–154. doi:10.1016/j.jmp.2008.12.005

Nouretdinov, I., Costafreda, S. G., Gammerman, A., Chervonenkis, A., Vovk, V., Vapnik, V., & Fu, C. H. (2011). Machine learning classification with confidence: application of transductive conformal predictors to MRI-based diagnostic and prognostic markers in depression. *Neuroimage, 56*(2), 809–813. doi:10.1016/j.neuroimage.2010.05.023

Patel, M. J., Andreescu, C., Price, J. C., Edelman, K. L., Reynolds, C. F., 3rd, & Aizenstein, H. J. (2015). Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. *International Journal of Geriatric Psychiatry, 30*(10), 1056–1067. doi:10.1002/gps.4262

Patel, M. J., Khalaf, A., & Aizenstein, H. J. (2016). Studying depression using imaging and machine learning methods. *Neuroimage: Clinical, 10*, 115–123. doi:10.1016/j.nicl.2015.11.003

Paulus, M. P., Huys, Q. J., & Maia, T. V. (2016). A Roadmap for the Development of Applied Computational Psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging,*

chinaXiv:201908.00074v1

1(5), 386–392. doi:10.1016/j.bpsc.2016.05.001

Pe, M. L., Vandekerckhove, J., & Kuppens, P. (2013). A diffusion model account of the relationship between the emotional flanker task and rumination and depression. *Emotion, 13*(4), 739–747. doi:10.1037/a0031628

Pizzagalli, D. A. (2014). Depression, stress, and anhedonia: Toward a synthesis and integrated model. *Annual Review of Clinical Psychology, 10*(1), 393–423. doi:10.1146/annurev-clinpsy-050212-185606

Radenbach, C., Reiter, A. M., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H. J., . . . Schlagenhauf, F. (2015). The interaction of acute and chronic stress impairs model-based behavioral control. *Psychoneuroendocrinology, 53*, 268–280. doi:10.1016/j.psyneuen.2014.12.017

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59–108. doi:10.1037/0033-295x.85.2.59

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873–922. doi:10.1162/neco.2008.12-06-420

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences, 20*(4), 260–281. doi:10.1016/j.tics.2016.01.007

Razi, A., Kahan, J., Rees, G., & Friston, K. J. (2015). Construct validation of a DCM for resting state fMRI. *Neuroimage, 106*, 1–14. doi:10.1016/j.neuroimage.2014.11.027

Robinson, O. J., & Chase, H. W. (2017). Learning and Choice in Mood Disorders: Searching for the Computational Parameters of Anhedonia. *Computational Psychiatry, 1*(1), 208–233. doi:10.1162/CPSY_a_00009

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: A systematic review and meta-analysis. *Psychological Medicine, 44*(10), 2029–2040. doi:10.1017/S0033291713002535

Rothkirch, M., Tonn, J., Kohler, S., & Sterzer, P. (2017). Neural mechanisms of reinforcement learning in unmedicated patients with major depressive disorder. *Brain, 140*(4), 1147–1157. doi:10.1093/brain/awx025

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., . . . Fava, M. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. *American Journal of Psychiatry, 163*(11), 1905–1917. doi:10.1176/ajp.2006.163.11.1905

Russo, S. J., & Nestler, E. J. (2013). The brain reward circuitry in mood disorders. *Nature Reviews Neuroscience, 14*(9), 609–625. doi:10.1038/nrn3381

Rutledge, R. B., Moutoussis, M., Smittenaar, P., Zeidman, P., Taylor, T., Hrynkiewicz, L., . . . Dolan, R. J. (2017). Association of Neural and Emotional Impacts of Reward Prediction Errors With Major Depression. *JAMA Psychiatry, 74*(8), 790–797. doi:10.1001/jamapsychiatry.2017.1713

Schnack, H. G., & Kahn, R. S. (2016). Detecting Neuroimaging Biomarkers for Psychiatric Disorders: Sample Size Matters. *Frontiers in Psychiatry, 7*, 50. doi:10.3389/fpsyt.2016.00050

Schwabe, L. (2013). Stress and the engagement of multiple memory systems: integration of animal and human studies. *Hippocampus, 23*(11), 1035–1043. doi:10.1002/hipo.22175

Schwabe, L., & Wolf, O. T. (2011). Stress-induced modulation of instrumental behavior: From goal-directed to habitual control of action. *Behavioural Brain Research, 219*(2), 321–328. doi:10.1016/j.bbr.2010.12.038

Sejnowski, T. J., Koch, C., & Churchland, P. S. (1988). Computational neuroscience. *Science, 241*(4871), 1299–1306. doi:10.1126/science.3045969

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*: Cambridge university press.

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine, 49*(9), 1426–1448. doi:10.1017/S0033291719000151

Singh, A., Thakur, N., & Sharma, A. (2016). *A review of supervised machine learning algorithms.* Paper presented at the International Conference on Computing for Sustainable Global Development, New Delhi, India.

Stephan, K. E., Iglesias, S., Heinzle, J., & Diaconescu, A. O. (2015). Translational Perspectives for Computational Neuroimaging. *Neuron, 87*(4), 716–732. doi:10.1016/j.neuron.2015.07.008

Stephan, K. E., Kasper, L., Harrison, L. M., Daunizeau, J., den Ouden, H. E., Breakspear, M., & Friston, K. J. (2008). Nonlinear dynamic causal models for fMRI. *Neuroimage, 42*(2), 649–662. doi:10.1016/j.neuroimage.2008.04.262

Stephan, K. E., & Mathys, C. (2014). Computational approaches to psychiatry. *Current Opinion in Neurobiology, 25*, 85–92. doi:10.1016/j.conb.2013.12.007

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage, 49*(4), 3099–3109. doi:10.1016/j.neuroimage.2009.11.015

Stephan, K. E., Schlagenhauf, F., Huys, Q. J. M., Raman, S., Aponte, E. A., Brodersen, K. H., . . . Heinz, A. (2017). Computational neuroimaging strategies for single patient predictions. *Neuroimage, 145*(Pt B), 180–199. doi:10.1016/j.neuroimage.2016.06.038

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., . . . Corlett, P. R. (2018). The Predictive Coding Account of Psychosis. *Biological Psychiatry, 84*(9), 634–643. doi:10.1016/j.biopsych.2018.05.015

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*: MIT press.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*: MIT press.

Tran, B. X., Vu, G. T., Ha, G. H., Vuong, Q. H., Ho, M. T., Vuong, T. T., . . . Ho, R. C. M. (2019). Global Evolution of Research in Artificial Intelligence in Health and Medicine: A Bibliometric Study. *Journal of Clinical Medicine, 8*(3), 360. doi:10.3390/jcm8030360

Turner, B. M., van Maanen, L., & Forstmann, B. U. (2015). Informing cognitive abstractions through neuroimaging: the neural drift diffusion model. *Psychological Review, 122*(2), 312–336. doi:10.1037/a0038894

van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: EZ, fast-dm, and DMAT. *Journal of Mathematical Psychology, 53*(6), 463–473. doi:10.1016/j.jmp.2009.09.004

Vos, T., Allen, C., Arora, M., Barber, R. M., Bhutta, Z. A., Brown, A., . . . Murray, C. J. L. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet, 388*(10053), 1545–1602. doi:10.1016/s0140-6736(16)31678-6

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology, 60*(6), 385–402. doi:10.1027/1618-3169/a000218

Wagstaff, K., Cardie, C., Rogers, S., & Schrödl, S. (2001). *Constrained k-means clustering with*

chinaXiv:201908.00074v1

background knowledge. Paper presented at the Proceedings of the Eighteenth International Conference on Machine Learning, Williamstown, MA, USA.

Wang, X. W., Nie, D., & Lu, B. L. (2014). Emotional state classification from EEG data using machine learning approach. *Neurocomputing, 129*, 94–106. doi:10.1016/j.neucom.2013.06.046

White, C. N., Ratcliff, R., Vasey, M., & McKoon, G. (2009). Dysphoria and memory for emotional material: A diffusion-model analysis. *Cognition & Emotion, 23*(1), 181–205. doi:10.1080/02699930801976770

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010a). Anxiety enhances threat processing without competition among multiple inputs: A diffusion model analysis. *Emotion, 10*(5), 662–677. doi:10.1037/a0019474

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010b). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology, 54*(1), 39–52. doi:10.1016/j.jmp.2010.01.004

WHO. (2018). Depression. Retrieved from http://www.who.int/news-room/fact-sheets/detail/depression

Wiecki, T. V., Poland, J., & Frank, M. J. (2015). Model-Based Cognitive Neuroscience Approaches to Computational Psychiatry: Clustering and Classification. *Clinical Psychological Science, 3*(3), 378–399. doi:10.1177/2167702614565359

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics, 7*, 14. doi:10.3389/fninf.2013.00014

Williams, L. M., Korgaonkar, M. S., Song, Y. C., Paton, R., Eagles, S., Goldstein-Piekarski, A., . . . Etkin, A. (2015). Amygdala Reactivity to Emotional Faces in the Prediction of General and Medication-Specific Responses to Antidepressant Treatment in the Randomized iSPOT-D Trial. *Neuropsychopharmacology, 40*(10), 2398–2408. doi:10.1038/npp.2015.89

Wunderlich, K., Dayan, P., & Dolan, R. J. (2012). Mapping value based planning and extensively trained choice in the human brain. *Nature Neuroscience, 15*(5), 786–791. doi:10.1038/nn.3068

chinaXiv:201908.00074v1

Computational psychiatry: A new perspective on research and clinical applications in depression

OU Jianxin[1]; WU Yin[1]; LIU Jinting[1]; LI Hong[1, 2, 3]

([1] *School of Psychology, Shenzhen University*; [2] *Shenzhen Key Laboratory of Affective and Social Cognitive Science;* [3] *Shenzhen Institute of Neuroscience, Shenzhen 518060, China*)

Abstract: Depression, a complex and heterogeneous mental disorder, leads to great global burdens of disease. Although diagnosis based on nosology is broadly used in several domains, it is still unable to direct the exploration of pathological mechanism of depression. In addition, several treatments developed by this diagnosis have poor outcomes due to its low prediction validity. Computational approaches to psychiatry remedy those limitations and help to improve understanding, prediction and treatment for depression by two complementary approaches: data-driven and theory-driven. Theory-driven approaches apply models to multiple levels of analysis from the prior knowledge or hypothesis of depression. Data-driven approaches, however, adopt machine-learning methods to analyze high-dimensional data to improve the diagnostic and predictive accuracies of depression, and eventually, promote the treatment effects. With the development and combination of these two approaches as well as the integration of resources, it is promising to cure depression and prevent it from occurrence.

Keywords: depression; computational psychiatry; computational models; machine learning; diagnosis; treatment

chinaXiv:201908.00074v1